Lecture 3: Clusters, Covariates, Compliance

Raymond Duch
Official Fellow Nuffield College
Director CESS Nuffield/Santiago

Centre for Experimental Social Sciences

Road Map to Lecture 3

- *Cluster random assignment*
- *Covariates*
- *Compliance*

Cluster Random Assignment

- Thus far we have focussed on individual-level treatments
- Individuals are often embedded in clusters where they receive either the treatment or control
    - Media market
    - Voting district or precinct
    - Classroom
- May be unavoidable or the level at which the intervention realistically takes place
- If potential outcomes differ across clusters it will lead to imprecise estimates

Centre for Experimental Social Sciences

PO: High Sampling Variability

		Classroom-level		*Cluster-level mean*	
School	Classroom	Y_i^c	Y_i^t	Y_i^c	Y_i^t
A	A-1	0	4		
A	A-2	1	5	1	5
A	A-3	2	6		
B	B-1	2	6		
B	B-2	3	7	3	7
B	B-3	4	8		
C	C-1	3	7		
C	C-2	4	8	4	8
C	C-3	5	9		
D	D-1	7	11		
D	D-2	8	12	8	12
D	D-3	9	13		

Centre for Experimental Social Sciences

PO: Low Sampling Variability

		Classroom-level		*Cluster-level mean*	
School	Classroom	Y_i^c	Y_i^t	Y_i^c	Y_i^t
A	A-1	0	4		
A	A-2	3	7	4	8
A	A-3	9	13		
B	B-1	2	6		
B	B-2	3	7	4	8
B	B-3	7	11		
C	C-1	1	5		
C	C-2	4	8	3.3	7.3
C	C-3	5	9		
D	D-1	4	8		
D	D-2	8	12	4.7	8.7
D	D-3	2	6		

Centre for Experimental Social Sciences

Cluster Random Assignment

- When clusters are the same size the ATE can be estimated the usual way
  - 4 in both cases
- Different standard errors
    - High variability = 2.9
    - Complete randomization = 1.6
    - Low variability = 0.57
- Penalty associated with clustering depends on the variability of the cluster-level means

SE with Clustered Design

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{k-1}\left\{\frac{mVar(\bar{Y}_j^c)}{N-m} + \frac{(N-m)Var(\bar{Y}_j^t)}{m} + 2Cov(Y_j^c, Y_j^t)\right\}}$$

- Clusters with very similar mean Yjc and mean Yjt would lead to smaller variances and lead to a more precise estimate of the ATE

Cluster Random Assignment

- Likely stuck with highly variable clusters based on
    - Geography
    - Institutions
    - Age groups
- Improving precision
    - Increase the number of clusters
    - Increasing the number of subjects per cluster will not have much of an effect on between cluster variance
    - Include covariates
    - sampclus in Stata allows you to play with the number of clusters and cluster size necessary to achieve adequate power

Covariate Adjustment?

- *Random assignment ensures unbiased estimation of the ATE*
  - *Omitted variables is addressed by random assignment*
  - *Including controls is not required*

Covariates Useful

- *Rescale dependent variable*
    - *Change from pre-test to post-test (diff-in-diff)*
    - *Potential outcomes have less variance*
- *Include in a regression analysis*
    - *Eliminate observed differences between treatment and control group*
    - *Reduce variability in outcomes*
    - *Results in more precise estimate of the treatment effect*
- *Check randomization process*
- *Construct blocks*

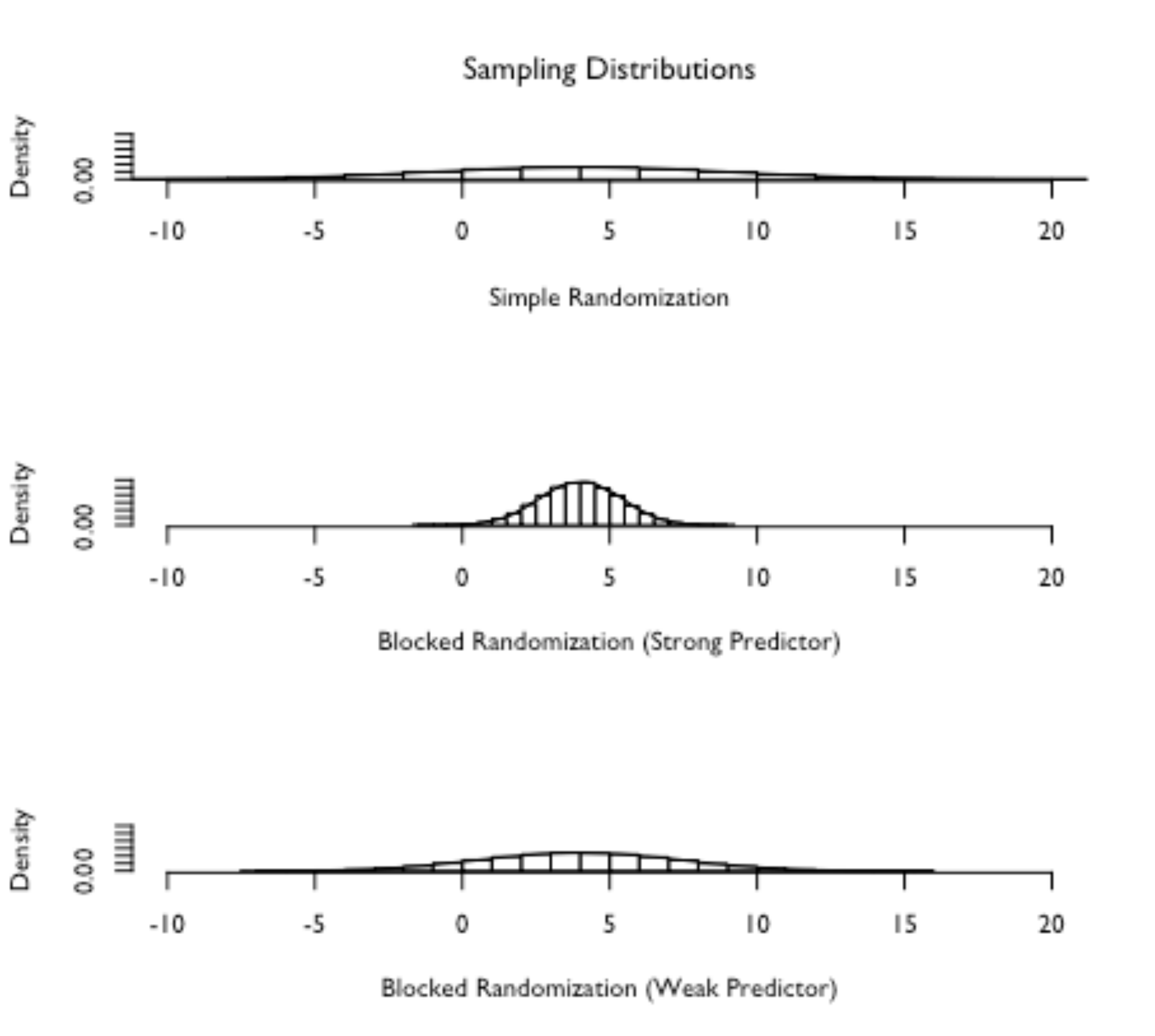
Covariate Rescaling

- *Pre-treatment covariates*
  - *Fixed constants that are observed prior to random assignment*
  - *Unaffected by treatment assignment*
- *Concerns*
  - *Budget constraints*
  - *Pre-test changes the way participants respond to the treatment*
    - *Violates excludability assumption*

Difference in Difference

$$
\begin{aligned}
E(\widehat{\mathrm{ATE}}) &= E[Y_i - X_i | D_i = 1] - E[Y_i - X_i | D_i = 0] \\
&= E[Y_i | D_i = 1] - E[X_i | D_i = 1] - E[Y_i | D_i = 0] + E[X_i | D_i = 0] \\
&= E[Y_i(1)] - E[Y_i(0)]. \qquad (4.2)
\end{aligned}
$$

Observation	y1	y0	D	x	x_weak
1	5	5	0	6	25
2	15	5	1	8	12
3	12	6	1	5	25
4	19	9	0	13	27
5	17	10	0	9	10
6	18	11	0	15	24
7	24	12	0	16	21
8	11	13	0	17	25
9	16	14	0	19	35
10	25	19	1	23	28
11	18	20	1	28	41
12	21	20	0	28	38
13	17	20	0	9	30
14	24	21	1	16	20
15	27	24	1	23	24
16	26	25	0	15	26
17	30	27	1	23	22
18	37	27	0	33	34
19	43	30	1	42	37
20	39	32	0	31	21
21	36	32	0	29	40
22	27	32	0	28	34
23	33	32	1	35	36
24	37	35	1	28	37
25	48	35	0	41	48
26	39	37	1	37	46
27	42	38	1	32	25
28	37	38	1	37	21
29	53	41	0	36	19
30	50	42	1	44	44
31	51	43	1	48	50
32	43	44	1	43	48
33	55	45	1	55	46
34	49	47	0	53	47
35	48	48	0	51	47
36	52	51	1	43	39
37	59	52	0	57	50
38	52	52	0	51	46
39	55	57	1	49	54
40	63	62	1	55	42



Sampling Distributions
Density
0.00
-10
-5
0
5
10
15
20
Simple Randomization
Density
0.00
-10
-5
0
5
10
15
20
Blocked Randomization (Strong Predictor)
Density
0.00
-10
-5
0
5
10
15
20
Blocked Randomization (Weak Predictor)

Covariates Using Regression

$$\begin{aligned} Y_i^* = Y_i - X_i &= Y_i(0)(1 - d_i) + Y_i(1)d_i - X_i \\ &= a + bd_i + u_i - X_i = a + bd_i + u_i^*. \end{aligned} \tag{4.8}$$

$$Y_i = Y_i(0)(1 - d_i) + Y_i(1)d_i = a + bd_i + cX_i + (u_i - cX_i), \tag{4.9}$$

Advantages of Covariates

- Can include several covariates as right-hand-side variables
- Reduces disturbance variability more effectively than rescaling
- Weakly predictive covariates do not reduce sampling variability
- Which pre-treatment covariates?
    - Previous research
    - Pilot testing
    - Theoretical intuition

Caution

- Experimental outcomes should not be used to decide which covariates are include
  - i.e., running several regressions and choosing the one that makes the ATE look best
- Plan in advance of the experiment
- Present difference-in-means and covariate-adjusted estimates

Covariate Imbalance

- Randomization may create imbalance in some covariates
  - Controlling for covariates reestablishes balance
- Imbalance also may alert problems with randomization
  - Retrace randomization process
  - Check with third parties executing the randomization
  - Correct possible errors

Balance Test (Panagopoulos et al. 2014)

Experimental conditions	*N*	Voted (Nov 2008)	Voted (Nov 2006)	Voted (Nov 2004)	Age (years)	Male	Partisan
Self + community high	1,000	64.4	29.2	43.1	27.5	39.9	80.4
Self + community low	1,000	66.8	30.2	44.4	28.0	38.4	83.2
Self only	1,000	64.0	26.9	40.0	28.0	40.1	78.9
Community high only	1,000	64.7	29.3	42.6	26.8	41.1	81.0
Community low only	1,000	64.7	27.8	41.3	27.0	41.9	80.3
Control	13,482	64.7	27.8	42.7	27.5	41.7	81.2
$p > F^{a}$		.83	.45	.43	.55	.31	.24

Figures in columns represent mean percentages unless otherwise indicated

[a] Test statistics generated using one way ANOVA to evaluate whether mean turnout levels differ across categories of random assignment. In all cases, we cannot reject the hypothesis of equal means at standard significance levels ($p < .05$), implying balance across groups

Block Randomisation

- Previous discussion demonstrated that blocking can improve precision
- It is possible to block on several variables simultaneously
- Biggest improvements in precision come from variables that strongly predict the outcome
    - Hard to know before experiment has been run
    - Look to previous studies for clues
- Blocking and covariate adjustment yield similar results with large sample sizes
    - Treatment and control groups of more than 100 subjects

One-sided Noncompliance

- Failure-to-treat
  - Individuals in the treatment group do not receive the treatment
  - May be due to miscommunication, manpower shortages, transportation problems, difficult to reach subjects, refused treatment

One-sided Noncompliance

- Compliance refers to those who were assigned the treatment and received it
- Noncompliance refers to those who were assigned the treatment but did not receive it or when those in the control group receive the treatment inadvertently
- ``One-sided" is the case when those assigned the treatment do not receive it

One-sided Noncompliance

- Cannot simply compare those actually treated with the control group
- Those treated are a ***non-random subset*** of the treatment original group
- Groups formed after random assignment cannot be expected to have the same potential outcomes which may lead to bias

Canvasing Example

- Those that have moved may still be listed on voter rolls
- Under random assignment they are equally likely to be assigned the treatment and control
- Canvassers will not reach those in the treatment but the controls will never be contacted
- Turnout will likely be lower among movers
- Excluding the movers from the treatment group will inflate the effect of canvassing
    - Comparing a control group made up of movers and non-movers and a treatment group made up of just non-movers

Expand Potential Outcomes

$$Y_i = d_i Y_i^t - (1 - d_i) Y_i^c$$

- Does not reflect that assigned treatment may not coincide with actual treatment. Let z denote the assigned treatment (0,1) and di(z) denote whether subject i was treated when assigned z.

GROUP	D(0)	D(1)
COMPILERS	0	1
NEVER-TAKERS	0	0

Centre for Experimental Social Sciences

Causal Effect of Assignment to the Treatment Group: Intent-To-Treat Effect

$$ITT_{i,D} = (d_i(z=1) - d_i(z=0))$$

$$ITT_D = \frac{1}{N}\sum_{i=1}^{N}(d_i(z=1)) = \alpha$$

$$ITT_{i,Y} = (Y_i(z=1) - Y_i(z=0))$$

$$ITT_Y = \frac{1}{N}\sum_{i=1}^{N}(Y_i(z=1) - Y_i(z=0))$$

$$\widehat{ITT} = (\bar{Y}(z=1) - \bar{Y}(z=0))$$

With 100% compliance *ATE* = *ITT* since *d* = *z*

Centre for Experimental Social Sciences

Intent-to-Treat-Effect

- Causal Effect of Assignment to the Treatment Group
- ITT captures the average effect of being assigned to the treatment group regardless of the proportion of the treatment group that was actually treated
- Measure of the overall success of an intervention in terms of changed outcomes
  - Did my program make an overall difference?
  - Noncompliance irrelevant

Complier Average Causal Effect

$$\text{CACE} = \frac{\text{ITT}}{\sigma}$$

- where σ is the share of those assigned to the treatment group receiving treatment
- CACE also referred to as Local Average Treatment Effect (LATE) and Treatment on Treated (TOT)
- ATE among Compliers
- CACE may be quite different from ATE
- Those who do not comply may have very different treatment effects

ITT, ATE: Potential Outcomes

Obs	$Y_i(0)$	$Y_i(1)$	$D_i = 0$	$D_i = 1$	Type
1	4	6	0	1	Complier
2	2	8	0	0	Never-Taker
3	1	5	0	1	Complier
4	5	7	0	1	Complier
5	6	10	0	1	Complier
6	2	10	0	0	Never-Taker
7	6	9	0	1	Complier
8	2	5	0	1	Complier
9	5	9	0	0	Never-Taker

Centre for Experimental Social Sciences

Compare ATT, ATE, and CACE

- ATE does not consider noncompliance:

$$\text{ATE} = \frac{2+6+4+2+4+8+3+3+4}{9} = 4$$

- ITT accounts for the fact that never-takers will not receive the treatment:

$$\text{ITT} = \frac{2+0+4+2+4+0+3+3+0}{9} = 2$$

- CACE is based on the subset of Compilers:

$$\text{CACE} = \frac{2+4+2+4+3+3}{6} = 3$$

Centre for Experimental Social Sciences

New Haven Voter Mobilization

- GOTV weeks leading up to the 1998 general election
- 7,090 registered voters living in one-voter households assigned to treatment and control
- Control group received no contact from the campaign
- Treatment group was visited by canvassers stressing the importance of voting
- After the campaign researchers checked public voting records to see who casted a ballot

New Haven Voter Mobilization

Turnout Rate	Treatment Group	Control Group
Among those contacted	54.43 (395)	
Among those not contacted	36.48 (1050)	37.54 (5645)
Overall	41.38 (1445)	37.45 (5645)

- ITT = 41.38 – 37.54 = 3.84
- σ= 395/1445 =0.273
- CACE=ITT/σ= 3.84/0.273 = 14.1

OLS, ITT and CACE

- Parameters of interest can be easily estimated using regression
- ITT using OLS
- Equivalent to difference in means
- CACE using two-stage squares
- Code and data are at \href{http://isps.yale.edu/FEDAI}{http://isps.yale.edu/FEDAI}

Centre for Experimental Social Sciences

New Haven Voter Mobilization: Estimating ITT

$$VOTED_i = \beta_0 + \beta_1\ ASSIGNED_i + \mu_i$$

	Estimate	Estimate	t-value	Pr(>|t|)
Intercept	0.3754	0.0065	58.23	0.0000
ASSIGNED	0.0385	0.0144	2.66	0.0079

Estimate for ASSIGNED matches what we got by hand

New Haven Voter Mobilization: Estimating α

$$TREATED_i = \gamma_0 + \gamma_1 \, ASSIGNED_i + e_i$$

	Estimate	Estimate	t-value	Pr(>|t|)
Intercept	0.00	0.00	24.89	0.00
ASSIGNED	0.27	0.01	23.30	0.00

As we found before, ~ 27% of the assigned treatment group received the treatment

New Haven Voter Mobilization: Estimating CACE

$$VOTED_i = \beta_0 + \beta_1 \, ASSIGNED_i + \mu_i$$

$$TREATED_i = \gamma_0 + \gamma_1 \, ASSIGNED_i + e_i$$

	Estimate	Estimate	t-value	Pr(>|t|)
Intercept	0.38	0.01	58.23	0.00
TREATED	0.14	0.05	2.68	0.01

- 2SLS model where TREATED is instrumented by ASSIGNED
  - It may be difficult to treat individuals due to reasons correlated with u_i
  - ASSIGNED is randomly determined so uncorrelated with u_i and related to TREATED
- Estimate for TREATED matches what we got by hand

Using Regression to Estimate ITT

Assumptions

- Non-interference (individuals unaffected by the treatment of another individual)
    - Someone made excited to vote by the canvasser may tell a neighbor in the control group
- Excludability (only the treatment exerts the effect)
    - Canvassers may say things other than what is contained in the script

Assumptions

- Anticipation of Noncompliance
    - The precision of the CACE estimator worsens as the rate of noncompliance goes up
    - Large samples required to offset high rates of noncompliance
- Placebo design
    - Researchers attempt to contact individuals assigned to receive the treatment
    - Those reached are then randomly allocated to two different groups
        - Treatment group
        - Placebo group receiving a ``non-treatment"

Placebo Design

- Nickerson (2008) canvassing experiment
    - GOTV (treatment)
    - Recycling (placebo)
- CACE estimated by comparing the outcomes for those in the treatment group to those in the placebo group
    - Random sample of Compliers who's untreated potential outcomes can be measured

Nickerson 2008

	Denver		Minneapolis		Pooled	
	Direct	Secondary	Direct	Secondary	Direct	Secondary
Percent Voting in GOTV Group	47.7% (3.0)	42.4% (2.9)	27.1% (3.1)	23.6% (3.0)		
Percent Voting in Recycling Group	39.1% (2.9)	36.9% (2.9)	16.2% (2.7)	17.3% (2.7)		
Estimated Treatment Effect	**8.6% (4.2)**	**5.5% (4.1)**	**10.9% (4.1)**	**6.4% (4.1)**	**9.8% (2.9)**	**6.0% (2.9)**
P-Value	0.02	0.09	<0.01	0.06	<0.01	0.02

Note. Numbers in parentheses represent standard errors. P-values test the one-tailed hypothesis. Pooled estimates are weighted averages of results for both cities.

Centre for Experimental Social Sciences

Placebo Design

- CACE estimated by comparing the outcomes for those in the treatment group to those in the placebo group
- Random sample of Compliers who's untreated potential outcomes can be measured
- Logic is that placebo design screens out Never-Takers
- Compliers in the treatment group are compared directly to Compliers in the untreated group
- Reduces noise from Never-Takers in both treatment and control groups
- Moves us to a world of ``full compliance"

Centre for Experimental Social Sciences

Placebo Design

- Downside is that not all Compliers receive the treatment
- Resources are wasted on those receiving the placebo
- Opportunity to collaborate with someone studying an unrelated topic

Centre for Experimental Social Sciences

Placebo Design

- The placebo and conventional design both allow estimation of the CACE
- Choice depends on the budget and compliance rate
- Under a fixed budget, the conventional design is preferable if the compliance rate is greater than 50%
- Canvassing studies often have a lower rate
- A pilot study may give a better idea of the expected compliance rate

Centre for Experimental Social Sciences

Nickerson 2008